REMARKS

This is in response to the Office Action dated February 12, 2008. The Office Action first reports that the Restriction Requirement has been withdrawn and that claims 15-29 are still pending in the application. The Applicant respectfully thanks the Examiner for his reconsideration and reinstatement of the claims. However, it is believed the Examiner made an error in that claims 30-34 would also be reinstated for the same reason since claims 30-34 are also simply dependent claims from independent claims 15. Hence, these claims are also listed above. If the undersigned is in error, please advise in the next communication.

The Office Action next reports that claims 15-29 recite non-statutory subject matter. In response, Applicant has amended claims 15-29 (and claims 30-34) to recite a "computer readable storage medium", which is believed preferred under U.S. practice. Support for this language is found at least in the specification at page 8, line 1 through page 9, line 5.

The Office Action next reports that claims 1-29 are rejected under 35 U.S.C. 103(a) as being unpatentable over Conki et al. (U.S. Patent 7,266,497) in view of Tzirkel-Hancock (U.S. Patent 6,275,795).

Referring first to claim 1, this claim recites:

> A method of ascertaining phoneme speech unit boundaries of adjacent speech units in speech data, the method comprising:
>
> receiving training data of speech waveforms with known boundary locations of phoneme speech units contained therein;
>
> processing the speech waveforms to obtain multi-frame acoustic feature pseudo-triphone representations of a plurality of pseudo-triphones in the speech data, each pseudo-triphone comprising a boundary location, a first phoneme speech unit preceding the boundary location and a second phoneme speech unit following the boundary location;
>
> clustering the multi-frame acoustic feature pseudo-triphone

representations as a function of acoustic similarity in a plurality of clusters;

training a refining model for each cluster;

receiving a second set of data of speech waveforms with initial boundary locations of adjacent phoneme speech units contained therein;

identifying pseudo-triphones in the second set of data and corresponding refining models for each of the pseudo-triphones; and

using the refining model for each corresponding pseudo-triphone for the second set of data to locate a new boundary location different than the initial boundary and provide output indicating the new boundary locations.

Conki et al. was cited as disclosing the last four steps of training, receiving a second set of data, identifying pseudo-triphones with corrected boundaries and outputting the same, while Tzirkel-Hancock was cited for disclosing the first three steps of receiving training data, processing the speech waveforms to obtain multi-frame acoustic feature pseudo-triphone representations and clustering the multi-frame acoustic feature pseudo-triphone representations as a function of acoustic similarity in a plurality of clusters.

Applicant respectfully traverses this rejection on at least two grounds that being Tzirkel-Hancock does not teach, suggest or render obvious the steps it is being cited for, and secondly, neither reference teaches, suggest or renders obvious the use of pseudo-triphones as presently recited.

Tzirkel-Hancock provides a system and method for determining articulation information from the speech signal of a speaker. Figure 5 provides an apparatus to extract such information and it is the operation of this apparatus that is contended to meet the first three steps of claim 1. At least an overview of the operation of

Figure 5 is found at col. 7, line 43 – col. 8, line 36, which is provided below for convenience.

> FIG. 5 shows an overview of an apparatus, according to a first embodiment of the present invention, used to extract the articulatory information from the speech signal and to produce a representation of the probable positions of the articulators based on the input speech signal. In this embodiment, the system is designed to output the result incrementally, i.e. to generate the output as soon as the input is received with a small delay.
>
> In FIG. 5, 41 is a preprocessor for extracting formant related information from the input speech signal, 42 is a circular buffer for storing signals from preprocessor 41; 43 is a segmenter that determines segment boundaries in the speech signals within circular buffer 42; 45 is an acoustic classifier for classifying each segment; 47 is a feature estimator for estimating the probability that the articulatory features have some value, e.g. open or closed; 49 represents a decision making block which estimates the values of the articulatory features; 51 is a memory which stores information needed for the segmenter 43, the acoustic classifier 45, the feature estimator 47 and the decision making block 49; 53 is a display for displaying, for example, the mid-sagittal plane of the human vocal system showing the articulators as the speaker speaks; and 55 represents a peripheral device, for example a printer or hard disc for storing the information which has been extracted.
>
> In operation, preprocessor 41 divides the input speech signal as it arrives into frames and determines, for each frame, signals representative of formant information of the input speech signal within that frame. For convenience, the signals that represent each frame of input speech will be referred to as "vectors". The remainder of the apparatus is used to estimate, for each vector, the articulatory feature values of interest, which requires some knowledge of the acoustic context of the vector. In this embodiment, only the local context of the vector is used, i.e. the input speech signal for neighbouring frames. This context information is determined by the operation of buffer 42, segmenter 43 and acoustic classifier 45 as follows: as each vector is determined it is fed into circular buffer 42 that buffers a small number of such vectors. As each new vector enters buffer 42, segmenter 43 decides whether the vector belongs to the current segment or whether the vector should start a new segment based on information derived by the apparatus during a first training session. If segmenter 43 decides that a new segment should be started then the buffer address of the first and last vector in the current segment are supplied to acoustic classifier 45.
>
> Once a segment, comprising a number of vectors, has been identified it is classified into one of a number of acoustic classes by acoustic classifier 45, each acoustic class being defined beforehand from training data from a second training session.

> The probable value of the articulatory features of interest are then calculated for each vector in the current segment, in feature extractor 47 and decision block 49 using the acoustic classification information obtained by acoustic classifier 45 and using models obtained beforehand during a third training session. This information may then be printed out directly for analysis or may be used to generate a corresponding image of the articulatory structures of the speaker as he/she speaks on display 53. (Emphasis Added)

Claim 1 recites, in part, a step comprising "clustering the multi-frame acoustic feature pseudo-triphone representations as a function of acoustic similarity in a plurality of clusters." Nowhere in passage provided above with respect to operation of the apparatus of Figure 5 is such a step found. Although the apparatus of Tzirkel-Hancock includes an acoustic classifier 45, this component does not cluster multi-frame acoustic feature pseudo-triphone representations as a function of acoustic similarity. Rather, as indicated above, the acoustic classifier identifies a segment as belonging to a class in order to obtain information that is used to obtain values of articulatory features. Neither does Conkie et al. disclose this feature.

As indicated above, it is submitted that neither Tzirkel-Hancock nor Conkie et al. teach, suggest or render obvious the use of pseudo-triphone representations of a plurality of pseudo-triphones in speech data, where each pseudo-triphone comprisesg a boundary location, a first phoneme speech unit preceding the boundary location and a second phoneme speech unit following the boundary location, as presently recited. The Office Action contends that HMM meets this definition, but it is respectfully submitted that HMM modeling by itself is not organized based on context dependency of on speech units, but rather a manner in which to model elements such as speech units based on ordered states. As stated in the Specification at page 17, lines 18-23, HMM is one form in which to model the pseudo-triphones, while other include Neural Networks (NN) or Gaussian Mixture

Models (GMM). By itself none of these models provide the recited context dependency of on speech units relative to the boundaries between speech units.

In view of the foregoing, Applicant respectfully requests withdrawal of the rejection and allowance of claim 1. Independent claim 15 recites similar features as those recited in claim 1, and thus, for at least the reasons discussed above, this claim is also believed allowable.

Dependent claims 2-14 and 16-34 each depend from claims 1 and 15, respectively. Each of these claims recite further features, which when combined with the features recited by their corresponding independent claim, and any intervening claims, recite further separately patentable inventions. In particular, many of these claims recite further features related to clustering, which as explained above is not taught by the cited references. Therefore, the further features related to clustering are clearly separately patentable.

The foregoing remarks are intended to assist the Office in examining the application and in the course of explanation may employ shortened or more specific or variant descriptions of some of the claim language. Such descriptions are not intended to limit the scope of the claims; the actual claim language should be considered in each case. Furthermore, the remarks are not to be considered exhaustive of the facets of the invention which are rendered patentable, being only examples of certain advantageous features and differences, which applicant's attorney chooses to mention at this time. For the foregoing reasons, applicant reserves the right to submit additional evidence showing the distinction between applicant's invention to be unobvious in view of the prior art.

Furthermore, in commenting on the references and in order to facilitate a better understanding of the differences that are expressed in the claims, certain

details of distinction between the same and the present invention have been mentioned, even though such differences do not appear in all of the claims. It is not intended by mentioning any such unclaimed distinctions to create any implied limitations in the claims.

An extension of time for consideration of this response is hereby requested. An online charge authorization for the extension of time fee is provided.

In view of the foregoing, Applicants respectfully request reconsideration of the application as amended. Withdrawal of the rejections and allowance of the pending claims is solicited.

The Director is authorized to charge any fee deficiency required by this paper or credit any overpayment to Deposit Account No. 23-1123.

Respectfully submitted,

WESTMAN, CHAMPLIN & KELLY, P.A.

By: /Steven M. Koehler, Reg. No. 36,188/
Steven M. Koehler, Reg. No. 36,188
900 Second Avenue South, Suite 1400
Minneapolis, Minnesota 55402
Phone: (612) 334-3222 Fax: (612) 334-3312

SMK:dkm